

# SCALING LIMITS OF MARKOV BRANCHING TREES AND GALTON-WATSON TREES CONDITIONED ON THE NUMBER OF VERTICES WITH OUT-DEGREE IN A GIVEN SET

DOUGLAS RIZZOLO

**ABSTRACT.** We generalize recent results of Haas and Miermont in [7] to obtain scaling limits of Markov branching trees whose size is specified by the number of nodes whose out-degree lies in a given set. We then show that this implies that the scaling limit of finite variance Galton-Watson trees conditioned on the number of nodes whose out-degree lies in a given set is the Brownian continuum random tree. The key to this is a generalization of the classical Otter-Dwass formula.

## 1. INTRODUCTION

Recently there has been considerable interest in the literature in studying the asymptotic properties of random trees. Much of the focus has been on limits of trees satisfying either nice consistency relations between trees of various sizes ([8]) or having a nice encoding as continuous functions on  $[0, \infty)$  (see [9] for an overview). Particular interest has been focused on limits of Galton-Watson trees conditioned on the total number of vertices. The standard techniques for proving the convergence of Galton-Watson trees to continuum trees are intimately connected with contour processes of these trees. However, conditioning on a subset of the vertices produces significantly more complicated contour processes than conditioning on the number of vertices so we must take a different approach. Techniques for handling Markov branching trees whose size is their number of leaves were developed in [7] and we will generalize this approach to Markov branching trees whose size is their number of vertices whose out-degree falls in a given set.

Our main result, the notation for which will be fully defined later, is the following theorem.

**Theorem 1.** *Let  $T$  be a critical Galton-Watson tree with offspring distribution  $\xi$  such that  $0 < \sigma^2 = \text{Var}(\xi) < \infty$  and let  $A \subseteq \{0, 1, 2, \dots\}$  contain 0. Suppose that for sufficiently large  $n$  the probability that  $T$  has exactly  $n$  vertices with out-degree in  $A$  is positive, and for such  $n$  let  $T_n^A$  be  $T$  conditioned to have exactly  $n$  vertices with out-degree in  $A$ , considered as a rooted unordered tree with edge lengths 1 and the uniform probability distribution  $\mu_{\partial_A T_n^A}$  on its vertices with out-degree in  $A$ . Then*

$$\frac{1}{\sqrt{n}} T_n^A \xrightarrow{d} \frac{2}{\sigma \sqrt{\xi(A)}} T_{1/2, \nu_2},$$

where the convergence is with respect to the rooted Gromov-Hausdorff-Prokhorov topology and  $T_{1/2, \nu_2}$  is the Brownian continuum random tree.

---

This material is based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No. DGE 1106400.

In the case  $A = \mathbb{Z}^+ = \{0, 1, 2, \dots\}$  we recover the classical result about the scaling limit of a Galton-Watson tree conditioned on its number of vertices first obtained in [1]. For other choices of  $A$  the result appears to be new. The condition that for sufficiently large  $n$  the probability that  $T$  has exactly  $n$  vertices with out-degree in  $A$  is positive is purely technical and could be dispensed with at the cost of chasing a constant (which may appear in the limit) through our computations. In addition to generalizing the results of [7], the key to proving this theorem is a generalization of the classical Otter-Dwass formula, which we prove in Section 3.1. The Otter-Dwass formula (see [12]) has been an essential tool in several proofs that the Brownian continuum random tree is the scaling limit of Galton-Watson trees conditioned on their number of vertices, including the original proof in [1] as well as newer proofs in [10] and [7]. While we follow the approach in [7], our generalization of Otter-Dwass formula should allow for proofs along the lines of [1] and [10] as well. Furthermore, with our results here, it should be straight forward to prove the analogous theorem in the infinite variance case using the approach in [7].

This paper is organized as follows. Section 2 introduces our basic notation, as well as the Markov branching trees and continuum trees that will arise for us as scaling limits. It concludes with our generalization of the scaling limits in [7]. Section 3 is devoted to our study of Galton-Watson trees. We begin by proving our generalization of the Otter-Dwass formula and we then use this to analyze the asymptotics of the partition at the root of a Galton-Watson tree. Bringing this all together, we finish with the proof of Theorem 1.

**Acknowledgment.** I would like to thank my adviser, Jim Pitman, for suggesting the problem that led to this paper and for helpful conversations and feedback along the way.

## 2. MODELS OF TREES

**2.1. Basic notation.** Fix a countably infinite set  $S$ ; we will consider the vertex sets of all graphs discussed to be subsets of  $S$ . A rooted ordered tree is a finite acyclic graph  $t$  with a distinguished vertex called the root and such that, if  $v$  is a vertex of  $t$ , the set of vertices in  $t$  that are both adjacent to  $v$  and further from the root than  $v$  with respect to the graph metric is linearly ordered. Let  $\mathcal{T}^S$  be the set of all rooted ordered trees whose vertex set is a subset of  $S$ . For  $t, s \in \mathcal{T}^S$ , define  $t \sim s$  if and only if there is a root and order preserving isomorphism from  $t$  to  $s$  and let  $\mathcal{T} = \mathcal{T}^S / \sim$  be the set of rooted ordered trees considered up to root and order preserving isomorphism. By a similar construction, we let  $\mathcal{T}^u$  be the set of rooted unordered trees considered up to root preserving isomorphism. If  $t$  is in  $\mathcal{T}$  or  $\mathcal{T}^u$  and  $v \in t$  is a vertex, the out-degree of  $v$  is the number of vertices in  $t$  that are both adjacent to  $v$  and further from the root than  $v$  with respect to the graph metric. The out-degree of  $v$  will simply be denoted by  $\deg(v)$ , since we will only ever discuss out-degrees. Fix a set  $A \subseteq \mathbb{Z}^+$  such that  $0 \in A$  and for  $t$  in  $\mathcal{T}$  or  $\mathcal{T}^u$  define  $\#_A t$  to be the number of vertices in  $t$  whose out-degree is in  $A$ . Furthermore, we define  $\mathcal{T}_{A,n}$  and  $\mathcal{T}_{A,n}^u$  by

$$\mathcal{T}_{A,n} = \{t \in \mathcal{T} : \#_A t = n\} \quad \text{and} \quad \mathcal{T}_{A,n}^u = \{t \in \mathcal{T}^u : \#_A t = n\}.$$

**2.2. Markov branching trees.** In this section we extend the notion of Markov branching trees developed in [7], where Markov branching trees were constructed separately in the cases  $A = \{0\}$  and  $A = \mathbb{Z}^+$ . Here we give a construction for general  $A$  such that  $0 \in A$ . Let  $\bar{\mathcal{P}}_n$

be the set of partitions of  $n$  and, for  $\lambda \in \bar{\mathcal{P}}_n$ , let  $p(\lambda)$  be the number of nonzero blocks in  $\lambda$  and  $m_j(\lambda)$  the number of blocks in  $\lambda$  equal to  $j$ . For convenience, we take  $\bar{\mathcal{P}}_1 = \{\emptyset, (1)\}$  and define  $p(\emptyset) = -1$ . Define  $\bar{\mathcal{P}}_1^A = \bar{\mathcal{P}}_1$  and for  $n \geq 2$ , define  $\bar{\mathcal{P}}_n^A$  by

$$\bar{\mathcal{P}}_n^A = \{\lambda \in \bar{\mathcal{P}}_n : p(\lambda) \notin A\} \cup \{\lambda \in \bar{\mathcal{P}}_{n-1} : p(\lambda) \in A\}.$$

Let  $(n_k)$  be an increasing sequence of integers. A sequence  $(q_{n_k})_{k \geq 1}$ , such that  $q_{n_k}$  is a probability measure on  $\bar{\mathcal{P}}_{n_k}^A$ , is called *compatible* if for each  $k$ ,  $q_{n_k}$  is concentrated on partitions  $\lambda = (\lambda_1, \dots, \lambda_p)$  such that  $q_{\lambda_i}$  is defined for all  $i$ . Suppose further that  $q_1$  is defined,  $q_{n_k}((n_k)) < 1$  if  $1 \notin A$  and  $q_2((1)) = 1$  if  $1 \in A$ . Our goal is to construct a sequence of laws  $(\mathbf{P}_{n_k}^q)_{k=1}^\infty$  such that  $\mathbf{P}_{n_k}^q$  is a law on  $\mathcal{T}_{A, n_k}^u$  and such that the subtrees above a vertex are conditionally independent given the degree of that vertex.

Define  $\mathbf{P}_1^q$  to be the law of the path with a root attached to a leaf by a path with  $G$  edges where  $G = 0$  if  $1 \in A$  and has the geometric distribution  $\mathbb{P}(G = j) = q_1(\emptyset)(1 - q_1(\emptyset))^j$ ,  $j \geq 0$  if  $1 \notin A$ . For  $k \geq 2$ ,  $\mathbf{P}_{n_k}^q$  is defined as follows: Choose  $\Lambda \in \bar{\mathcal{P}}_{n_k}^A \setminus \{(n_k)\}$  according to  $q_{n_k}(\cdot | \bar{\mathcal{P}}_{n_k}^A \setminus \{(n_k)\})$  and independently choose  $G'$  with  $G' = 1$  if  $1 \in A$  and  $G'$  has a geometric distribution

$$\mathbb{P}(G' = j) = (1 - q_{n_k}((n_k)))q_{n_k}((n_k))^{j-1}, j \geq 1,$$

if  $1 \notin A$ . Let  $(T_1, T_2, \dots, T_{p(\Lambda)})$  be a vector of trees, independent of  $G'$ , such that the  $T_i$  are independent and  $T_i$  has distribution  $\mathbf{P}_{\Lambda_i}^q$  and let  $T$  be the tree that results from attaching the roots of the  $T_i$  to the same new vertex and then if  $G' = 1$  call this vertex the root, and otherwise attach that vertex to a new root by a path with  $G' - 1$  edges.  $\mathbf{P}_{n_k}^q$  is defined to be the law of  $T$ . An easy induction shows that  $\mathbf{P}_{n_k}^q$  is concentrated on the set of unordered rooted trees with exactly  $n_k$  vertices whose out-degree is in  $A$ .

To connect with [7], if  $(n_k) = (1, 2, 3, \dots)$ , the case  $A = \{0\}$  corresponds to the  $\mathbf{P}_n^q$  defined in [7] and the case  $A = \mathbb{Z}^+$  corresponds to the  $\mathbf{Q}_n^q$  defined in [7]. Other choices of  $A$  interpolate between these two extremes. A sequence  $(T_{n_k})_{k \geq 1}$  such that for each  $k$ ,  $T_{n_k}$  has law  $\mathbf{P}_{n_k}^q$  for some choice of  $A$  and  $q$  (independent of  $n$ ) is called a Markov branching family. For ease of notation, we will generally drop the subscript  $k$  and it will be implicit that we are only considering  $n$  for which the quantities discussed are defined.

**2.3. Trees as metric measure spaces.** The trees we have been talking about can naturally be considered as metric spaces with the graph metric. That is, the distance between vertices is the number of edges on the path connecting them. Let  $(t, d)$  be a tree equipped with the graph metric. For  $a > 0$ , we define  $at$  to be the metric space  $(t, ad)$ , i.e. the metric is scaled by  $a$ . This is equivalent to saying the edges have length  $a$  rather than length 1 in the definition of the graph metric. More, generally we can attach a positive length to each edge in  $t$  and use these in the definition of the graph metric. Moreover, the trees we are dealing with are rooted so we consider  $(t, d)$  as a pointed metric space with the root as the point. Moreover, we are concerned with the vertices whose out-degree is in  $A$ , so we attach a measure  $\mu_{\partial_A t}$ , which is the uniform probability measure on  $\partial_A t = \{v \in t : \deg(v) \in A\}$ . If we have a random tree  $T$ , this gives rise to a random pointed metric measure space  $(T, d, \text{root}, \mu_{\partial_A T})$ . To make this last concept rigorous, we need to put a topology on pointed metric measure spaces. This is hard to do in general, but note that the pointed metric measure spaces that come from the trees we are discussing are compact.

Let  $\mathcal{M}_w$  be the set of equivalence classes of compact pointed metric measure spaces (equivalence here being up to point and measure preserving isometry). It is worth pointing out that  $\mathcal{M}_w$  actually is a set in the sense of ZFC, though this takes some work to show. We metrize  $\mathcal{M}_w$  with the pointed Gromov-Hausdorff-Prokhorov metric (see [7]). Fix  $(X, d, \rho, \mu), (X', d', \rho, \mu') \in \mathcal{M}_w$  and define

$$d_{\text{GHP}}(X, X') = \inf_{(M, \delta)} \inf_{\substack{\phi: X \rightarrow M \\ \phi': X' \rightarrow M}} [\delta(\phi(\rho), \phi'(\rho')) \vee \delta_H(\phi(X), \phi'(X')) \vee \delta_P(\phi_*\mu, \phi'_*\mu')],$$

where the first infimum is over metric spaces  $(M, \delta)$ , the second infimum is over isometric embeddings  $\phi$  and  $\phi'$  of  $X$  and  $X'$  into  $M$ ,  $\delta_H$  is the Hausdorff distance on compact subsets of  $M$ , and  $\delta_P$  is the Prokhorov distance between the pushforward  $\phi_*\mu$  of  $\mu$  by  $\phi$  and the pushforward  $\phi'_*\mu'$  of  $\mu'$  by  $\phi'$ . Again, the definition of this metric has potential to run into set-theoretic difficulties, but they are not terribly difficult to resolve.

**Proposition 1** (Proposition 1 in [7]). *The space  $(\mathcal{M}_w, d_{\text{GHP}})$  is a complete separable metric space.*

An  $\mathbb{R}$ -tree is a complete metric space  $(T, d)$  with the following properties:

- For  $v, w \in T$ , there exists a unique isometry  $\phi_{v,w} : [0, d(v, w)]$  with  $\phi_{v,w}(0) = v$  to  $\phi_{v,w}(d(v, w)) = w$ .
- For every continuous injective function  $c : [0, 1] \rightarrow T$  such that  $c(0) = v$  and  $c(1) = w$ , we have  $c([0, 1]) = \phi_{v,w}([0, d(v, w)])$ .

If  $(T, d)$  is an  $\mathbb{R}$ -tree, every choice of root  $\rho \in T$  and probability measure  $\mu$  on  $T$  yields an element  $(T, d, \rho, \mu)$  of  $\mathcal{M}_w$ . With this choice of root also comes a height function  $\text{ht}(v) = d(v, \rho)$ . The leaves of  $T$  can then be defined as a point  $v \in T$  such that  $v$  is not in  $[[\rho, w[ := \phi_{\rho,w}([0, \text{ht}(w)))$  for any  $w \in T$ . The set of leaves is denoted  $\mathcal{L}(T)$ .

**Definition 1.** *A continuum tree is an  $\mathbb{R}$ -tree  $(T, d, \rho, \mu)$  with a choice of root and probability measure such that  $\mu$  is non-atomic,  $\mu(\mathcal{L}(T)) = 1$ , and for every non-leaf vertex  $w$ ,  $\mu\{v \in T : [[\rho, v] \cap [[\rho, w] = [[\rho, w]]\} > 0$ .*

The last condition says that there is a positive mass of leaves above every non-leaf vertex. We will usually just refer to a continuum tree  $T$ , leaving the metric, root, and measure as implicit. A continuum random tree (CRT) is an  $(\mathcal{M}_w, d_{\text{GHP}})$  valued random variable that is almost surely a continuum tree. The continuum random trees we will be interested in are those associated with self-similar fragmentation processes.

**2.4. Self-similar fragmentations.** For any set  $B$ , let  $\mathcal{P}_B$  be the set of countable partitions of  $B$ , i.e. countable collections of disjoint sets whose union is  $B$ . For  $n \in \overline{\mathbb{N}} := \mathbb{N} \cup \{\infty\}$ , let  $\mathcal{P}_n := \mathcal{P}_{[n]}$ . Suppose that  $\pi = (\pi_1, \pi_2, \dots) \in \mathcal{P}_n$  (here and throughout we index the blocks of  $\pi$  in increasing order of their least elements), and  $B \subseteq \overline{\mathbb{N}}$ . Define the restriction of  $\pi$  to  $B$ , denoted by  $\pi|_B$  or  $\pi \cap B$ , to be the partition of  $[n] \cap B$  whose elements are the blocks  $\pi_i \cap B$ ,  $i \geq 1$ . We topologize  $\mathcal{P}_n$  by the metric

$$d(\pi, \sigma) := \frac{1}{\inf\{i : \pi \cap [i] \neq \sigma \cap [i]\}}.$$

It is worth noting that this is, in fact, an ultra-metric, i.e. for  $\pi^1, \pi^2, \pi^3 \in \mathcal{P}_n$ , we have

$$d(\pi^1, \pi^2) \leq \max(d(\pi^1, \pi^3), d(\pi^2, \pi^3)).$$

Note that  $(\mathcal{P}_n, d)$  is compact for all  $n$ .

**Definition 2** (Definition 3.1 in [3]). *Consider two blocks  $B \subseteq B' \subseteq \mathbb{N}$ . Let  $\pi$  be a partition of  $B$  with  $\#\pi = n$  non-empty blocks ( $n = \infty$  is allowed), and  $\pi^{(\cdot)} = \{\pi^{(i)}, i = 1, \dots, n\}$  be a sequence in  $\mathcal{P}_{B'}$ . For every integer  $i$ , we consider the partition of the  $i$ -th block  $\pi_i$  of  $\pi$  induced by the  $i$ -th term  $\pi^{(i)}$  of the sequence  $\pi^{(\cdot)}$ , that is,*

$$\pi_{|\pi_i}^{(i)} = \left( \pi_j^{(i)} \cap \pi_i : j \in \mathbb{N} \right).$$

*As  $i$  varies in  $[n]$ , the collection  $\left\{ \pi_j^{(i)} \cap \pi_i : i, j \in \mathbb{N} \right\}$  forms a partition of  $B$ , which we denote by  $\text{Frag}(\pi, \pi^{(\cdot)})$  and call the fragmentation of  $\pi$  by  $\pi^{(\cdot)}$ .*

It is relatively easy to show that  $\text{Frag}$  is Lipschitz continuous in the first variable, and continuous in an appropriate sense in the second. It is also relatively easy to show that if  $\pi$  is an exchangeable partition and  $\pi^{(i)}$  is a sequence of independent exchangeable partitions (also independent of  $\pi$ ), then  $\pi$  and  $\text{Frag}(\pi, \pi^{(\cdot)})$  are jointly exchangeable. See chapter 3 of [3] for both of these facts. We will use the  $\text{Frag}$  function to define the transition kernels of our fragmentation processes.

Define

$$\mathcal{S}^\downarrow = \left\{ (s_1, s_2, \dots) : s_1 \geq s_2 \geq \dots \geq 0, \sum_{i=1}^{\infty} s_i \leq 1 \right\},$$

and

$$\mathcal{S}_1 = \left\{ (s_1, s_2, \dots) \in [0, 1]^{\mathbb{N}} \mid \sum_{i=1}^{\infty} s_i \leq 1 \right\},$$

and endow both with the topology they inherit as subsets of  $[0, 1]^{\mathbb{N}}$  with the product topology. Observe that  $\mathcal{S}^\downarrow$  and  $\mathcal{S}_1$  are compact. For a partition  $\pi \in \mathcal{P}_\infty$ , we define the asymptotic frequency  $|\pi_i|$  of the  $i$ 'th block by

$$|\pi_i| = \lim_{n \rightarrow \infty} \frac{|\pi_i \cap [n]|}{n},$$

provided this limit exists. If all of the blocks of  $\pi$  have asymptotic frequencies, we define  $|\pi| \in \mathcal{S}_1$  by  $|\pi| = (|\pi_1|, |\pi_2|, \dots)$ .

**Definition 3** (Definition 3.3 in [3]). *Let  $\Pi(t)$  be an exchangeable, càdlàg  $\mathcal{P}_\infty$ -valued process such that  $\Pi(0) = \mathbf{1}_{\mathbb{N}} := (\mathbb{N}, 0, \dots)$  such that*

- (1)  $\Pi(t)$  almost surely possesses asymptotic frequencies  $|\Pi(t)|$  simultaneously for all  $t \geq 0$  and
- (2) if we denote by  $B_i(t)$  the block of  $\Pi(t)$  which contains  $i$ , then the process  $t \mapsto |B_i(t)|$  has right-continuous paths.

We call  $\Pi$  a self-similar fragmentation process with index  $\alpha \in \mathbb{R}$  if and only if, for every  $t, t' \geq 0$ , the conditional distribution of  $\Pi(t + t')$  given  $\mathcal{F}_t$  is that of the law of  $\text{Frag}(\pi, \pi^{(\cdot)})$ , where  $\pi = \Pi(t)$  and  $\pi^{(\cdot)} = (\pi^{(i)}, i \in \mathbb{N})$  is a family of independent random partitions such that for  $i \in \mathbb{N}$ ,  $\pi^{(i)}$  has the same distribution as  $\Pi(t'|\pi_i|^\alpha)$ .

The existence of these processes is non-trivial, but they do in fact exist. This makes sense even for  $\alpha < 0$  and, in this case, one can show that for any  $t > 0$  the set of singletons of  $\Pi(t)$  has positive asymptotic frequency. See corollary 3.2 in [3].

One important tool for studying self-similar fragmentations is the equivalent of the Lévy-Itô decomposition of Lévy processes. Suppose, for the moment, that  $\Pi$  is a self-similar fragmentation process with  $\alpha = 0$  (these are also called homogeneous fragmentations). In this case, it turns out that  $\Pi$  is a Feller process as is  $\Pi_{|[n]}$  for every  $n$ . Thus it is natural to try to identify the jump rates of these processes. For  $\pi \in \mathcal{P}_n \setminus \{\mathbf{1}_{[n]}\}$ , let

$$q_\pi = \lim_{t \rightarrow 0^+} \mathbb{P}(\Pi_{|[n]}(t) = \pi).$$

By exchangeability, it is obvious that  $q_\pi = q_{\sigma(\pi)}$  for every permutation  $\sigma$  of  $[n]$ . Less obvious, but still true, is that the law of  $\Pi$  is determined by the jump rates  $\{q_\pi : \pi \in \mathcal{P}_n \setminus \{\mathbf{1}_{[n]}\}, n \in \mathbb{N}\}$ . Furthermore, there is a nice description of these rates. For  $\pi \in \mathcal{P}_n$  and  $n' \in \{n, n+1, \dots, \infty\}$ , define

$$\mathcal{P}_{n', \pi} = \{\pi' \in \mathcal{P}_{n'} : \pi'_{|[n]} = \pi\}.$$

**Proposition 2** (Propositions 3.2 and 3.3 in [3]). *Suppose we have a family  $\{q_\pi : \pi \in \mathcal{P}_n \setminus \{\mathbf{1}_{[n]}\}, n \in \mathbb{N}\}$ . This family is the family of jump rates of some homogeneous fragmentation  $\Pi$  if and only if there is an exchangeable measure  $\mu$  on  $\mathcal{P}_\infty$  satisfying*

- (1)  $\mu(\{\mathbf{1}_\mathbb{N}\}) = 0$  and,
- (2)  $\mu(\{\pi \in \mathcal{P}_\infty : \pi_{|[n]} \neq \mathbf{1}_{[n]}\}) < \infty$  for every  $n \geq 2$ ,

*such that  $\mu(\mathcal{P}_{\infty, \pi}) = q_\pi$ . Furthermore, this correspondence is bijective, and we call  $\mu$  the splitting rate of  $\Pi$ .*

We are after a Lévy-Itô decomposition of  $\mu$ . For  $n \in \mathbb{N}$ , let  $\varepsilon^{(n)}$  be the partition of  $\mathbb{N}$  with exactly two blocks,  $\{n\}$  and  $\mathbb{N} \setminus \{n\}$  and define

$$\epsilon = \sum_{n=1}^{\infty} \delta_{\varepsilon^{(n)}}.$$

For a measure  $\nu$  on  $\mathcal{S}^\downarrow$  such that  $\nu(\{\mathbf{1}\}) = 0$  and  $\int_{\mathcal{S}^\downarrow} (1 - s_1) \nu(ds) < \infty$ , define a measure  $\rho_\nu$  on  $\mathcal{P}_\infty$  by

$$\rho_\nu(\cdot) = \int_{\mathcal{S}^\downarrow} \rho_s(\cdot) \nu(ds).$$

**Theorem 2** (Theorem 3.1 in [3]). *Let  $\mu$  be the splitting rate of a homogeneous fragmentation. Then there exists a unique  $c \geq 0$  and a unique measure  $\nu$  on  $\mathcal{S}^\downarrow$  with  $\nu(\{\mathbf{1}\}) = 0$  and  $\int_{\mathcal{S}^\downarrow} (1 - s_1) \nu(ds) < \infty$ , such that*

$$\mu = c\epsilon + \rho_\nu.$$

The interpretation of this is that  $c$  is the erosion coefficient, i.e. the rate at which mass is lost continuously, and  $\nu$  is the dislocation measure, i.e. it measures the rate of macroscopic fragmentation. From this theorem, it is clear that every homogeneous fragmentation process is characterized by the pair  $(c, \nu)$ . Given a homogeneous fragmentation  $\Pi^0(t)$  with parameters  $(0, \nu)$ , and  $\alpha < 0$ , we can construct an  $\alpha$ -self-similar fragmentation with parameters  $(\alpha, 0, \nu)$  by a time change. Let  $\pi^i(t)$  be the block of  $\Pi^0$  that contains  $i$  at time  $t$  and define

$$T_i(t) = \inf \left\{ u \geq 0 : \int_0^u |\pi^i(r)|^{-\alpha} dr \right\}.$$

For  $t \geq 0$ , let  $\Pi(t)$  be the partition such that  $i, j$  are in the same block of  $\Pi(t)$  if and only if they are in the same block of  $\Pi^0(T_i(t))$ . Then  $(\Pi(t), t \geq 0)$  is a self-similar fragmentation with characteristics  $(\alpha, 0, \nu)$ . See [2] for details.

We will need trees associated to fragmentations with characteristics  $(\alpha, 0, \nu)$ , where  $\alpha < 0$  and  $\nu(\sum_i s_i < 1) = 0$ . What these assumptions tell us is that there is no continuous loss of mass due to erosion ( $c = 0$ ), mass is not lost during macroscopic fragmentations ( $\nu(\sum_i s_i < 1) = 0$ ), and the fragmentation eventually becomes the partition into singletons (Proposition 2 in [2], the rate of convergence is given in Proposition 14 in [5]). Henceforth, we let  $\Pi$  be such a self-similar fragmentation.

The tree associated to a fragmentation process  $\Pi$  is a continuum random tree that keeps track of when blocks split apart and the sizes of the resulting blocks. For a continuum tree  $(T, \mu)$  and  $t \geq 0$ , let  $T_1(t), T_2(t), \dots$  be the tree components of  $\{v \in T : \text{ht}(v) > t\}$ , ranked in decreasing order of  $\mu$ -mass. We call  $(T, \mu)$  self-similar with index  $\alpha < 0$  if for every  $t \geq 0$ , conditionally on  $(\mu(T_i(t)), i \geq 1)$ ,  $(T_i(t), i \geq 1)$  has the same law as  $(\mu(T_i(t))^{-\alpha} T^{(i)}(t), i \geq 1)$  where the  $T^{(i)}$ 's are independent copies of  $T$ .

The following summarizes the parts of Theorem 1 and Lemma 5 in [6] that we will need.

**Theorem 3.** *Let  $\Pi$  be a  $(\alpha, 0, \nu)$ -self-similar fragmentation with  $\alpha < 0$  and  $\nu$  as above and let  $F := |\Pi|^\downarrow$  be its ranked sequence of asymptotic frequencies. There exists an  $\alpha$ -self-similar CRT  $(T_{-\alpha, \nu}, \mu_{-\alpha, \nu})$  such that, writing  $F'(t)$  for the decreasing sequence of masses of the connected components of  $\{v \in T_{-\alpha, \nu} : \text{ht}(v) > t\}$ , the process  $(F'(t), t \geq 0)$  has the same law as  $F$ . Furthermore,  $T_F$  is a.s. compact.*

The choice of where to put negative signs in the notation in the above theorem is to conform with the notation of [7]. The Brownian CRT is the  $-1/2$ -self-similar random tree with dislocation measure  $\nu_2$  given by

$$\int_{\mathcal{S}^\downarrow} \nu_2(d\mathbf{s}) f(\mathbf{s}) = \int_{1/2}^1 \sqrt{\frac{2}{\pi s_1^3 (1-s_1)^3}} ds_1 f(s_1, 1-s_1, 0, 0, \dots).$$

Since we will always have  $c = 0$ , we will drop it and for a measure  $\nu$  satisfying the above conditions and  $\gamma > 0$ , we refer to  $(-\gamma, \nu)$  as fragmentation pair, which is associated to a  $(-\gamma, \nu)$ -self-similar fragmentation.

**2.5. Convergence of Markov branching trees.** We first recall some of the main results of [7]. Let  $A \subseteq \mathbb{Z}^+$  contain 0 and let  $(q_n)$  be a compatible sequence of probability measures satisfying the conditions of Section 2.2. Define  $\bar{q}_n$  to be the push forward of  $q_n$  onto  $\mathcal{S}^\downarrow$  by  $\lambda \mapsto \lambda / \sum_i \lambda_i$ .

**Theorem 4** (Theorems 1 and 2 in [7]). *Suppose that  $A = \{0\}$  or  $A = \mathbb{Z}^+$ . Further suppose that there is a fragmentation pair  $(-\gamma, \nu)$  with  $0 < \gamma < 1$  and a function  $\ell : (0, \infty) \rightarrow (0, \infty)$ , slowly varying at  $\infty$  (or  $\gamma = 1$  and  $\ell(n) \rightarrow 0$ ) such that, in the sense of weak convergence of finite measures on  $\mathcal{S}^\downarrow$ , we have*

$$n^\gamma \ell(n) (1 - s_1) \bar{q}_n(ds) \rightarrow (1 - s_1) \nu(ds).$$

*Let  $T_n$  have law  $\mathbf{P}_n^q$  and view  $T_n$  as a random element of  $\mathcal{M}_w$  with the graph distance and the uniform probability measure  $\mu_{\partial_A T_n}$  on  $\partial_A T_n = \{v \in T_n : \deg v \in A\}$ . Then we have the convergence in distribution*

$$\frac{1}{n^\gamma \ell(n)} T_n \rightarrow T_{\gamma, \nu},$$

*with respect to the rooted Gromov-Hausdorff-Prokhorov topology.*

The case where  $A = \{0\}$  this is a special case of Theorem 1 in [7] and the case  $A = \mathbb{Z}^+$  is Theorem 2 in the same paper. The case  $A = \mathbb{Z}^+$  is proved by reduction to the  $A = \{0\}$  case. We extend this to the case of general  $A$  containing 0.

**Theorem 5.** *The conclusions of Theorem 4 are valid if the only assumption on  $A \subseteq \mathbb{Z}^+$  is that  $0 \in A$ .*

As argued at the start of Section 4 in [7], we may assume that  $q_1(\emptyset) = 1$ . This is because each leaf is connected to the rest of the tree by a stalk of vertices with out-degree one and geometric length. Setting  $q_1(\emptyset) = 1$  collapses these to be length one. Since these stalks are independent from one another, with probability approaching one, this costs  $\log(n)$  in the Gromov-Hausdorff-Prokhorov metric. This is negligible since we are scaling by  $(n^\gamma \ell(n))^{-1}$ . Let  $t$  be a rooted unordered tree with  $n$  vertices whose out-degree is in  $A$  and let  $t^\circ$  be the tree obtained from  $t$  by attaching a leaf to every non-leaf vertex of  $t$  whose out-degree is in  $A$ .

Define the inclusion  $\iota : \bar{\mathcal{P}}_{n-1} \rightarrow \bar{\mathcal{P}}_n$  by  $\iota(\lambda) = (\lambda, 1)$ . We now define a sequence  $q_n^\circ$  of probability measures on  $\bar{\mathcal{P}}_n$ . Define  $q_1(\emptyset) = 1$  and for  $n \geq 2$ ,

$$q_n^\circ(\lambda) = \begin{cases} q_n(\lambda) & \text{if } \lambda \in \bar{\mathcal{P}}_n^A \setminus \iota(\bar{\mathcal{P}}_n^A \cap \bar{\mathcal{P}}_{n-1}), \\ q_n(\lambda) + q_n(\lambda') & \text{if } \lambda \in \bar{\mathcal{P}}_n^A \text{ and } \lambda = \iota(\lambda') \text{ for some } \lambda' \in \bar{\mathcal{P}}_n^A \cap \bar{\mathcal{P}}_{n-1}, \\ q_n(\lambda') & \text{if } \lambda \notin \bar{\mathcal{P}}_n^A \text{ and } \lambda = \iota(\lambda') \text{ for some } \lambda' \in \bar{\mathcal{P}}_n^A \cap \bar{\mathcal{P}}_{n-1}, \\ 0 & \text{otherwise.} \end{cases}$$

**Lemma 1.** *If  $T_n$  has distribution  $\mathbf{P}_n^q$  then  $T_n^\circ$  has distribution  $\mathbf{P}_n^{q^\circ}$ .*

*Proof.* We prove this by induction. The result is clear for  $n = 1$ . For  $n \geq 2$ , we condition on the root partition. Indeed, since in both  $T_n^\circ$  and a tree with law  $\mathbf{P}_n^{q^\circ}$  the subtrees attached to the root are independent given the root partition, by induction (and a little care about when the partition at the root is  $(n)$ ), we need only check that the laws of the partitions at the root agree. This, however, is immediate from the construction of  $q_n^\circ$ .  $\square$

Therefore Theorem 5 is an immediate consequence of the following lemma.



**Lemma 2.** *If*

$$n^\gamma \ell(n)(1 - s_1) \bar{q}_n(d\mathbf{s}) \rightarrow (1 - s_1) \nu(d\mathbf{s}),$$

*then*

$$n^\gamma \ell(n)(1 - s_1) \bar{q}_n^\circ(d\mathbf{s}) \rightarrow (1 - s_1) \nu(d\mathbf{s}).$$

*Proof.* Let  $f : \mathcal{S}^\downarrow \rightarrow \mathbb{R}$  be Lipschitz continuous (with respect to the uniform norm) with both the uniform norm and Lipschitz constant bounded by  $K$ . Observe that for  $\lambda \in \bar{\mathcal{P}}_n$ ,

$$\left| f\left(\frac{\iota(\lambda)}{n+1}\right) - f\left(\frac{\lambda}{n}\right) \right| \leq K \sum_{i=1}^{p(\lambda)} \frac{\lambda_i}{n(n+1)} + \frac{K}{n+1} = \frac{2K}{n+1}.$$

Letting  $g(\mathbf{s}) = (1 - s_1)f(\mathbf{s})$ , we have

$$\begin{aligned} |\bar{q}_n^\circ(g) - \bar{q}_n(g)| &\leq \sum_{\lambda \in \bar{\mathcal{P}}_n^A \cap \bar{\mathcal{P}}_{n-1}} q_n(\lambda) \left| \left(1 - \frac{\lambda_1}{n}\right) f\left(\frac{\iota(\lambda)}{n}\right) - \left(1 - \frac{\lambda_1}{n-1}\right) f\left(\frac{\lambda}{n-1}\right) \right| \\ &\leq \sum_{\lambda \in \bar{\mathcal{P}}_n^A \cap \bar{\mathcal{P}}_{n-1}} q_n(\lambda) \left( \frac{K\lambda_1}{n(n+1)} + \frac{2K}{n+1} \right) \\ &\leq \frac{3K}{n+1}. \end{aligned}$$

Multiplying by  $n^\gamma \ell(n)$ , we see that this upper bound goes to 0 and the result follows.  $\square$

*Proof of Theorem 5.* Note that, if  $a > 0$ , then  $d_{\text{GHP}}(at, at^\circ) \leq a$ . Consequently

$$d_{\text{GHP}}\left(\frac{1}{n^\gamma \ell(n)} T_n, \frac{1}{n^\gamma \ell(n)} T_n^\circ\right) \leq \frac{1}{n^\gamma \ell(n)} \rightarrow 0.$$

Since  $(n^\gamma \ell(n))^{-1} T_n^\circ \rightarrow T_{\gamma, \nu}$  by Lemma 2 and Theorem 4,  $(n^\gamma \ell(n))^{-1} T_n \rightarrow T_{\gamma, \nu}$  as well.  $\square$

### 3. GALTON-WATSON TREES

Let  $\xi = (\xi_i)_{i \geq 0}$  be a probability distribution with mean less than or equal to 1, and assume that  $\xi_1 < 1$ . A Galton-Watson tree with offspring distribution  $\xi$  is a random element  $T$  of  $\mathcal{T}$  with law

$$\text{GW}_\xi(t) = \mathbb{P}(T = t) = \prod_{v \in t} \deg(v).$$

The fact that  $\xi$  has mean less than or equal to 1 implies that the right hand side defines an honest probability distribution of  $\mathcal{T}$ .

**3.1. Otter-Dwass type formulae.** In this section we develop a transformation of rooted ordered trees that takes Galton-Watson trees to Galton-Watson trees. This transformation is motivated by the observation that the number of leaves in a Galton-Watson tree is distributed like the progeny of a Galton-Watson tree with a related offspring distribution. This simple observation was first made in [11]. Let  $\xi = (\xi_i)_{i \geq 0}$  be a probability distribution with mean

less than or equal to 1, and assume that  $\xi_1 < 1$ . Let  $T$  be a Galton-Watson tree with offspring distribution  $\xi$  and let

$$C(z) = \sum_{i=1}^{\infty} \mathbb{P}(\#_{\{0\}}T = i)z^i$$

be the probability generating function of the number of leaves of  $T$ . Furthermore, let

$$\phi(z) = \sum_{i=0}^{\infty} \xi_{i+1}z^i.$$

Decomposing by the root degree, we see that  $C(z)$  satisfies the functional equation

$$C(z) = \xi_0 z + C'(z)\phi(C(z)).$$

Solving for  $C(z)$  yields

$$(3.1) \quad C(z) = z \left( \frac{\xi_0}{1 - \phi(C(z))} \right).$$

Define

$$(3.2) \quad \theta(z) = \frac{\xi_0}{1 - \phi(z)}.$$

Observe that  $\theta$  has nonnegative coefficients,  $[z^0]\theta(z) = \xi_0/(1 - \xi_1)$  and  $\theta(1) = 1$ . Thus the coefficients of  $\theta$  are a probability distribution, call it  $\zeta = (\zeta_i)_{i \geq 0}$ .

**Proposition 3.** *Let  $T$  be a Galton-Watson tree with offspring distribution  $\xi$  and let  $T'$  be a Galton-Watson tree with offspring distribution  $\zeta$  where  $\xi$  and  $\zeta$  are related as above. Then for all  $k \geq 1$ ,  $\mathbb{P}(\#_{\{0\}}T = k) = \mathbb{P}(\#_{\mathbb{Z}^+}T' = k)$ . Also,  $T'$  is critical (subcritical) if and only if  $T$  is critical (subcritical).*

*Proof.* The computations above show that the probability generating functions for  $\#_{\{0\}}T$  and  $\#_{\mathbb{Z}^+}T'$  satisfy the same functional equation and the Lagrange inversion formula implies they have the same coefficients. The criticality claims follows from Equation (3.2), which can also be used to obtain higher moments of  $\zeta$ .  $\square$

**Corollary 1.** *Let  $\mathcal{F}_n$  be an ordered forest of  $n$  independent Galton-Watson trees all with offspring distribution  $\xi$ . Let  $\zeta$  be related to  $\xi$  as in Proposition 3. Let  $(X_i)_{i \geq 1}$  be an i.i.d. sequence of  $\zeta$  distributed random variables and let  $S_k = \sum_{i=1}^k (X_i - 1)$ . Let  $\#_{\{0\}}\mathcal{F}_n$  denote the number of leaves in  $\mathcal{F}_n$ . Then for  $1 \leq k \leq n$*

$$\mathbb{P}(\#_{\{0\}}\mathcal{F}_n = k) = \frac{k}{n} \mathbb{P}(S_k = -n).$$

*Proof.* This follows immediately from Proposition 3 and the Otter-Dwass formula (see [12]).  $\square$

This relationship between  $T$  and  $T'$  can also be proved in a more probabilistic fashion. Indeed, by taking a more probabilistic approach we can get a more general result that includes the results in [11] as a special case. To prove the result in full generality, it is more convenient to work with the depth-first queue of  $T$  than with  $T$  itself.

For  $x \in \mathbb{Z}^{\mathbb{N}} = \mathbb{Z}^{\{1,2,3,\dots\}}$ , let  $\tau_{-1} = \inf \{n : \sum_{i=1}^n x_i = -1\}$ . Let  $\mathcal{D}$  be the set of sequences of excursion increments in  $\mathbb{Z}^{\mathbb{N}}$  that are bounded below by  $-1$ . Formally,

$$\mathcal{D} = \{x \in \mathbb{Z}^{\mathbb{N}} : x_i \geq -1 \text{ for } i \geq 1, x_i = 0 \text{ for } i \geq \tau_{-1}(x), \tau_{-1}(x) < \infty\}.$$

For  $t \in \mathcal{T}$  with  $n$  vertices, index the vertices  $V$  of  $t$  from 1 to  $n$  by order of appearance on the depth-first walk of  $t$ . Define  $DQ(t) = (DQ_k(t))_{k=0}^{\infty}$  by  $DQ_k(t) = \deg v_k - 1$  for  $k \leq n$  and 0 for  $k > n$ , which are the increments of the depth-first queue of  $t$ . Note that  $DQ(t) \in \mathcal{D}$ . Furthermore  $t \mapsto DQ(t)$  is a bijection from  $\mathcal{T}$  to  $\mathcal{D}$  (see e.g. [12]).

Let  $\pi_n$  be the projection onto the  $n$ 'th coordinate of  $\mathbb{Z}^{\mathbb{N}}$  and let  $\mathcal{F}_n = \sigma(\pi_k, k \leq n)$ . Let  $\theta_n$  be the shift  $(\theta_n x)(i) = x(n+i)$ . Let  $N'$  be a stopping time with respect to  $(\mathcal{F}_n)$ . Let  $N^0 = 0$  and for  $i \geq 1$  define  $N^i = N^{i-1} + (N' \wedge \tau_{-1}) \circ \theta_{N^{i-1}}$ . Define  $\hat{x}$  by

$$\hat{x}(k) = \begin{cases} \sum_{i=N^{k-1}+1}^{N^k} x(i) & \text{if } N^k < \infty \\ 0 & \text{if } N^k = \infty. \end{cases}$$

**Proposition 4.** *If  $x \in \mathcal{D}$ , then  $\hat{x} \in \mathcal{D}$ .*

*Proof.* The only non-trivial part is to see that for each  $x \in \mathcal{D}$  there exists  $k$  such that  $N^k = \tau_{-1}$ . Clearly  $N^1 \leq \tau_{-1}$ . Let  $k = \max\{i : N^i \leq \tau_{-1}\}$ . Suppose, for the sake of contradiction, that  $N^k < \tau_{-1}$ . We then have that  $\sum_{i=1}^{N^k} x(i) \geq 0$ , so  $\sum_{i=N^{k+1}}^{\tau_{-1}} x(i) \leq -1$ , so  $N^{k+1} \leq \tau_{-1}$ , which is our contradiction.  $\square$

Combining these ideas, we obtain the following theorem.

**Theorem 6.** *Let  $\xi$  be a probability distribution on  $\mathbb{Z}_+$  with  $0 < \xi_0 < 1$ . Suppose that  $T$  is a Galton-Watson tree with offspring distribution  $\xi$ . Let  $N'$  be a stopping time and let  $\hat{T}$  be the tree determined by  $\widehat{DQ(T)}$  by the bijection above. Let  $X = (X_1, X_2, \dots)$  be a vector with i.i.d. entries distributed like  $P(X_1 = k) = \xi_{k+1}$ . Then  $\hat{T}$  is a Galton-Watson tree whose offspring distribution is the law of  $1 + \sum_{i=1}^{N^1(X)} X_i$ . Furthermore, if  $\mathbb{E}|X_1| < \infty$  and  $\mathbb{E}(N' \wedge \tau_{-1}) < \infty$  then*

$$\mathbb{E} \left( 1 + \sum_{i=1}^{N^1(X)} X_i \right) = 1 + \mathbb{E}X_1 \mathbb{E}(N' \wedge \tau_{-1}),$$

and if, additionally,  $\mathbb{E}X_1 = 0$  (i.e.  $T$  is critical) and  $\text{Var}(X_1) = \sigma^2 < \infty$ , then

$$\text{Var} \left( 1 + \sum_{i=1}^{N^1(X)} X_i \right) = \sigma^2 \mathbb{E}(N' \wedge \tau_{-1}).$$

*Proof.* Define  $R(X)$  to be the vector with  $R(X)_k = X_k \mathbf{1}(k \leq \tau_{-1}(X))$ . It is well known that  $DQ(T) =_d R(X)$  and that the vectors  $\{(X_{N^i+1}, \dots, X_{N^{i+1}})\}_{i=0}^{\infty}$  are i.i.d. Consequently  $\hat{X}$  is the vector of increments of a random walk with jump distribution given by the law of  $\sum_{i=1}^{N^1(X)} X_i$ . Observing that  $\widehat{DQ(T)} =_d \widehat{R(X)} = R(\hat{X})$  shows that  $\hat{T}$  is a Galton-Watson tree with the appropriate offspring distribution. The last claims follow from Wald's equations.  $\square$

Let us give a specific example of how the general theorem above may be applied. For a nonempty subset  $A$  of  $\mathbb{Z}^+$  let  $\mathcal{D}_n^A$  be the set of paths in  $\mathcal{D}$  with exactly  $n$  terms no later than  $\tau_{-1}(x)$  in  $A - 1$ . When  $A = \mathbb{Z}^+$ , we just write  $\mathcal{D}_n$ . Note that for every  $n \geq 1$ ,  $t \mapsto DQ(t)$  restricts to a bijection from  $\mathcal{T}_{A,n}$  to  $\mathcal{D}_n^A$ . We obtain the following plethora of Otter-Dwass type formulae.

**Corollary 2.** *Fix  $A \subseteq \mathbb{Z}^+$  such that  $0 \in A$  and define  $N'(x) = \inf\{i : x_i + 1 \in A\}$ . Define  $T$ ,  $\hat{T}$ , and  $X$  as in Theorem 6 and let  $\hat{X}_1, \hat{X}_2, \dots$  be i.i.d. distributed like  $1 + \sum_{i=1}^{N^1(X)} X_i$ . Then*

$$\mathbb{P}(\#_A T = n) = \mathbb{P}(\hat{T} = n) = \frac{1}{n} \mathbb{P}\left(\sum_{i=1}^n \hat{X}_i = -1\right).$$

*The corresponding result for forests also holds. Furthermore, if  $T$  is critical with variance  $0 < \sigma^2 < \infty$ , then  $\text{Var}(\hat{X}_i) = \sigma^2/\xi(A)$ .*

*Proof.* This follows from the observation that, with this  $N'$ ,  $x \in \mathcal{D}_n^A$  if and only if  $\hat{x} \in \mathcal{D}_n$ . The formula for the variance follows from the fact that  $N'$  is geometric with parameter  $\xi(A)$ .  $\square$

We note that, in the context of Corollary 2, the same construction can be done directly on the trees without first passing to the depth-first queue, though setting up the formalism for the proof and the proof itself are slightly more involved. The idea is a lifeline construction. You proceed around the tree in the order of the depth-first walk and when you encounter a vertex whose degree is in  $A$  you label the edges and vertices on the path from the vertex to the root that are not yet labeled by that vertex. This labeled path can be considered the lifeline of the vertex. A new tree is constructed by letting the root be first vertex encountered whose degree is in  $A$  and attaching vertices whose degree is in  $A$  to the earliest vertex whose lifeline touches its own. Going through the details of this helps make this transformation more concrete, so the case of  $A = \{0\}$  is included below.

Suppose that  $t \in \mathcal{T}_{\{0\},n}$  and label the leaves by the order they appear in the depth-first walk of  $t$ . We will now color all of the edges of  $t$ . Color every edge on the path from leaf 1 to the root with 1. Continuing in increasing order of their labels, color all edges on the path from leaf  $i$  to the root that are not colored with an element of  $\{1, 2, \dots, i-1\}$  with  $i$ , until all edges are colored. Note that for any  $1 \leq k \leq n$  the subtree spanned by leaves  $\{1, \dots, k\}$  is colored by  $\{1, \dots, k\}$  and an edge is colored by an element in  $\{1, \dots, k\}$  if and only if it is in this subtree. Furthermore, the path from leaf  $k$  to any edge colored  $k$  contains only edges colored  $k$ . See Figure 1 for an example of such a coloring. Call two edges of  $t$  coincident if they share a common vertex.

**Lemma 3.** *If  $t$  is colored as above and  $2 \leq j \leq n$ , then there is exactly one edge colored  $j$  that is coincident to an edge with a smaller color.*

*Proof.* First we show existence. Consider the path from leaf  $j$  to the root. Let  $e$  be the last edge this path that is not contained in the subtree spanned by leaves  $\{1, \dots, j-1\}$ . By construction this edge is colored  $j$  and is coincident to an edge colored by an element of  $\{1, \dots, j-1\}$ .

To see uniqueness, suppose that  $f$  is an edge with the desired properties. Then  $f$  is on the path from  $j$  to the root and  $f$  is coincident to an edge in the subtree spanned by leaves  $\{1, \dots, j-1\}$ . If  $f$  contains the root, then  $f$  is the last edge on the path from  $j$  to the root that is colored  $j$ , i.e.  $f = e$ . Otherwise, after  $f$ , we finish the path from  $j$  to the root within this subtree. Hence  $f$  is the last edge on the path from  $j$  to the root that is colored  $j$  and again  $f = e$ .  $\square$

With  $t$  labeled as above we define a rooted plane tree with  $n$  vertices, called the life-line tree and denoted  $\tilde{t}$ , as follows. The vertex set of  $\tilde{t}$  is  $\{1, 2, \dots, n\}$ , 1 is the root. Furthermore, if  $i < j$ ,  $i$  is adjacent to  $j$  if  $i$  is the smallest number such that there exist coincident edges  $e_1, e_2$  in  $t$  with  $e_1$  colored  $i$  and  $e_2$  colored  $j$ . Finally, the children of a vertex are ordered by the appearance of the corresponding leaves in the depth-first search of  $t$ . See Figure 1 for an example of this map.

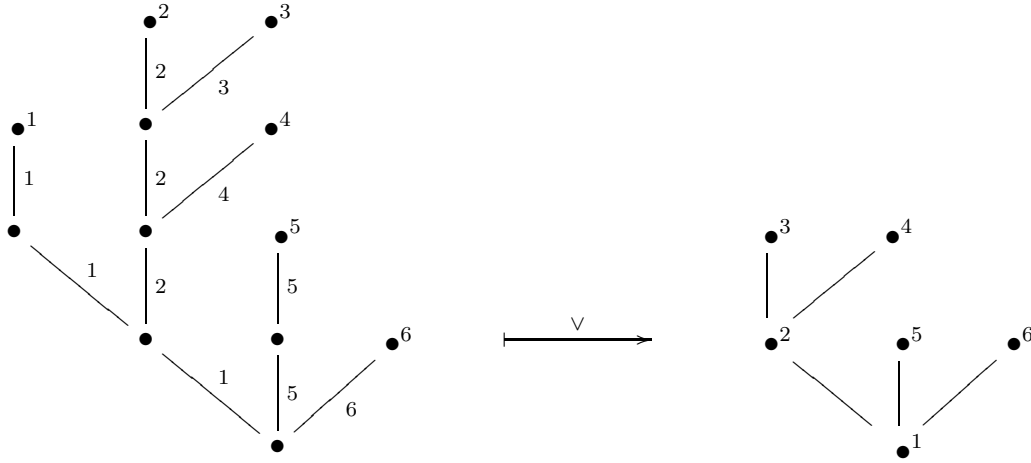


FIGURE 1. A colored tree and its image under  $V$

**Lemma 4.** *The life-line tree is a tree.*

*Proof.* We must show that  $\tilde{t}$  is connected and acyclic. Suppose that  $\tilde{t}$  has at least two components. Let  $j$  be the smallest vertex not in the same component as 1. By Lemma 3, there exists  $1 \leq i < j$  and coincident edges  $e_1, e_2$  in  $t$  labeled  $i$  and  $j$  respectively. Thus  $i$  is adjacent to  $j$ , a contradiction.

Suppose that  $\tilde{t}$  contains a cycle. Let  $j$  be the largest vertex in this cycle. Then  $j$  is adjacent to two smaller vertices, contradicting our definition of  $\tilde{t}$ .  $\square$

Let  $(v_1, \dots, v_k)$  be the list of vertices (ordered by order of appearance in the depth-first walk of  $t$ ) in  $t$  that are children of vertices on the path from 1 to the root, but not actually on that path themselves. Let  $t_{v_i}$  be the plane subtree of  $t$  above  $v_i$ . It is then easily verified that  $\tilde{t}$  is obtained by joining the trees  $(\tilde{t}_{v_1}, \dots, \tilde{t}_{v_k})$  to a common root with their natural order (and renaming the vertices as they appear in the depth first walk).

**Lemma 5.** *Let  $T$  be a Galton-Watson tree with offspring distribution  $\xi$ . Condition  $T$  on the event that the first leaf on the depth-first walk of  $T$  has height  $n$  and that there are exactly  $k$  vertices in  $T$  that are children of vertices on the path from the root to the first leaf on the depth-first walk that are not on this path themselves. Let  $v_1^n, \dots, v_k^n$  be these vertices (again in order of appearance) and let  $T_{v_j^n}$  be the plane subtree of  $T$  above  $v_j^n$ . The collection  $\{T_{v_j^n}\}_{j=1}^k$  is a collection of i.i.d Galton-Watson trees with common distribution  $T$ .*

*Proof.* Let  $t_1, \dots, t_k$  be rooted ordered trees. Let  $A$  be the set of trees  $t$  such that the first leaf on the depth-first walk of  $t$  has height  $n$  and that there are exactly  $k$  vertices  $v_1, \dots, v_k$  (listed in order of appearance on the depth-first walk) in  $t$  that are children of vertices on the path from the root to the first leaf on the depth-first walk that are not on this path themselves. Let  $B \subseteq A$  be the set of trees such that  $t_j$  is the tree above  $v_j$  for all  $1 \leq j \leq k$ . Let  $C$  be the set sequences that appear as the sequence of degrees of vertices on the path from the root to the left-most leaf of a tree in  $A$ . Note that

$$\mathbb{P}(T \in A) = \sum_{(y_1, \dots, y_{n+1}) \in C} \prod_{i=1}^{n+1} \xi_{y_i}$$

since, given  $(y_1, \dots, y_{n+1}) \in C$ , there are  $k$  places to attach trees to the path from the root to the left-most leaf, and we sum over all ways of doing this. Consequently, we have

$$\begin{aligned} \mathbb{P}(T_{v_j^n} = t_j, 1 \leq j \leq k) &= \frac{1}{\mathbb{P}(A)} \sum_{t \in B} \mathbb{P}(T = t) \\ &= \frac{1}{\mathbb{P}(A)} \sum_{(y_1, \dots, y_{n+1}) \in C} \prod_{i=1}^{n+1} \xi_{y_i} \prod_{i=1}^k \mathbb{P}(T = t_i) \\ &= \prod_{i=1}^k \mathbb{P}(T = t_i). \end{aligned}$$

□

**Theorem 7.** *Let  $T$  be a Galton-Watson tree with offspring distribution  $\xi$  and define  $\tilde{T}$  and  $T'$  as above (see Proposition 3 for  $T'$ ). Then  $\tilde{T} \stackrel{d}{=} T'$ .*

*Proof.* Let  $t$  be a rooted plane tree and consider  $\mathbb{P}(\tilde{T} = t)$ . If  $t$  has one vertex, it is clear that  $\mathbb{P}(\tilde{T} = t) = \mathbb{P}(T' = t)$ . Suppose that the result is true for all trees with less than  $n$  vertices, and suppose that  $t$  has  $n$  vertices. Let  $t_1, \dots, t_k$  be the subtrees of  $t$  attached to the root of  $t$ , listed in order of appearance of the depth-first walk of  $t$ . Let  $A_{i,k}$  be the event that the first leaf to appear on the depth-first walk of  $T$  has height  $i$  and that there are  $k$  vertices in  $T$  that are children of vertices on the path from the root to the first leaf on the depth-first walk that are not on this path themselves. Let  $v_1^i, \dots, v_k^i$  be these vertices (again in order of appearance) and let  $T_{v_j^i}$  be the plane subtree of  $T$  above  $v_j^i$ . Lemma 5 shows that, for fixed  $i$ , conditionally on  $A_{i,k}$ , the  $T_{v_j^i}$  are i.i.d. distributed like  $T$ . From our discussion above, we have that conditionally on  $A_{i,k}$ ,  $\tilde{T} = t$  if and only if  $\tilde{T}_{v_j^i} = t_j$  for all  $j$ . Since  $t_j$  has fewer

than  $n$  vertices, the inductive hypothesis implies

$$\begin{aligned}\mathbb{P}(\check{T} = t) &= \sum_{i=1}^{\infty} \mathbb{P}(T \in A_{i,k}) \mathbb{P}(\check{T}_{v_j^i} = t_j \text{ for all } j \mid T \in A_i) \\ &= \left( \prod_{j=1}^k \mathbb{P}(T' = t_j) \right) \left( \sum_{i=1}^{\infty} \mathbb{P}(T \in A_{i,k}) \right).\end{aligned}$$

Hence it remains to show that

$$\sum_{i=1}^{\infty} \mathbb{P}(T \in A_{i,k}) = \zeta_k.$$

Let  $(X_i)_{i=0}^{\infty}$  be i.i.d. distributed like  $\xi$ . We then have

$$\mathbb{P}(T \in A_{i,k}) = \mathbb{P}\left(X_0 = 0, \sum_{j=1}^i (X_j - 1) = k, X_j - 1 \geq 0 \text{ for } 1 \leq j \leq i\right) = \xi_0 [z^k] \phi(z)^i,$$

where for a power series  $\psi(z)$ ,  $[z^k]\psi$  is the coefficient of  $z^k$ . Thus we have

$$\sum_{i=1}^{\infty} \mathbb{P}(T \in A_{i,k}) = \xi_0 [z]^k \sum_{i=0}^{\infty} \phi(z)^i = [z^k] \frac{\xi_0}{1 - \phi(z)} = \zeta_k,$$

where the interchange of limits is justified by positivity of the coefficients involved and we may start the second sum at 0 since  $k \geq 1$ .  $\square$

**3.2. The partition at the root.** Let  $\xi = (\xi_i)_{i \geq 0}$  be a probability distribution with mean 1 and variance  $0 < \sigma_1^2 < \infty$ . Let  $T$  be a Galton-Watson tree with offspring distribution  $\xi$  (denote the law of  $T$  by  $\text{GW}_{\xi}$ ). Let  $A \subseteq \mathbb{Z}^+$  contain 0 and construct  $\hat{T}$  as in Corollary 2. Then, by Theorem 6,  $\hat{T}$  is a Galton-Watson tree. Let  $\zeta$  be its offspring distribution. Again by Theorem 6,  $\zeta$  has mean 1 and variance  $\sigma^2 = \sigma_1^2 / \xi(A)$ . Assume that for sufficiently large  $n$ ,  $\mathbb{P}(\#_A T = n) > 0$ . Let  $T_n^A$  be  $T$  conditioned to have exactly  $n$  vertices with out-degree in  $A$  (whenever this conditioning makes sense).

For a  $t$  be rooted unordered tree with exactly  $n$  vertices with out-degree in  $A$ , let  $\Pi^A(t)$  be the partition of  $n$  or  $n - 1$  (depending on whether or not the degree of the root of  $t$  is in  $A$ ) defined by the number of vertices with out-degree in  $A$  in the subtrees of  $t$  attached to the root.

**Lemma 6.** (i) Considered as an unordered tree, the law of  $T_n^A$  is equal to  $\mathbf{P}_n^q$  where, for  $n \geq 2$  such that  $T_n^A$  is defined, and  $\lambda = (\lambda_1, \dots, \lambda_p) \in \bar{\mathcal{P}}_n^A$ , we have

$$q_n(\lambda) = \mathbb{P}(\Pi^A(T_n^A) = \lambda) = \frac{p!}{\prod_{j \geq 1} m_j(\lambda)!} \xi(p) \frac{\prod_{i=1}^p \mathbb{P}(\#_A T = \lambda_i)}{\mathbb{P}(\#_A T = n)}.$$

(ii) Let  $X_1, X_2, \dots$  be i.i.d. distributed like  $\mathbb{P}(\#_A T = n)$  and  $\tau_k = X_1 + \dots + X_k$ . We have

$$\mathbb{P}(p(\Pi^A(T_n^A)) = p) = \xi(p) \frac{\mathbb{P}(\tau_p = n - \mathbf{1}(p \in A))}{\mathbb{P}(\tau_1 = n)},$$

and  $\mathbb{P}(\Pi^A(T_n^A) \in \cdot \mid \{p(\Pi^A(T_n^A)) = p\})$  is the law of a non-increasing rearrangement of  $(X_1, \dots, X_p)$  conditionally on  $X_1 + \dots + X_p = n - \mathbf{1}(p \in A)$ .

*Proof.* (i) Letting  $c_\emptyset(T_n^A)$  be the root degree of  $T_n^A$  and  $a_1, \dots, a_p \in \mathbb{N}$  with sum  $n - \mathbf{1}(p \in A)$  we have

$$(3.3) \quad \mathbb{P}(c_\emptyset(T_n^A) = p, \#_A[(T_n^A)_i] = a_i, 1 \leq i \leq p) = \xi(p) \frac{\prod_{i=1}^p \mathbb{P}(\#_A T = a_i)}{\mathbb{P}(\#_A T = n)}.$$

Part (i) now follows by considering the number of sequences  $(a_1, \dots, a_p)$  with the same decreasing rearrangement.

(ii) This follows from Equation (3.3).  $\square$

To simplify notation, let  $q_n$  be the law of  $\Pi^A(T_n^A)$  and let  $\mathbf{1}_p = \mathbf{1}(p \in A)$ . Let  $(S_r, r \geq 0)$  be a random walk with step distribution  $(\zeta_{i+1}, i \geq -1)$ . By Corollary 2, we have

$$(3.4) \quad q_n(p(\lambda) = p) = \xi(p) \frac{\frac{p}{n - \mathbf{1}_p} \mathbb{P}(S_{n - \mathbf{1}_p} = -p)}{\frac{1}{n} \mathbb{P}(S_n = -1)} = \frac{n}{n - \mathbf{1}_p} \hat{\xi}(p) \frac{\mathbb{P}(S_{n - \mathbf{1}_p} = -p)}{\mathbb{P}(S_n = -1)},$$

where  $\hat{\xi}(p) = p\xi(p)$  is the size-biased distribution of  $\xi$ .

Define  $\bar{q}_n$  to be the pushforward of  $q_n$  onto  $\mathcal{S}^\downarrow$  by the map  $\lambda \mapsto \lambda / \sum_i \lambda_i$ .

For a sequence  $(x_1, x_2, \dots)$  of non-negative numbers such that  $\sum_i x_i < \infty$ , let  $i^*$  be a random variable with

$$\mathbb{P}(i^* = i) = \frac{x_i}{\sum_{j \geq 1} x_j}.$$

The random variable  $x_1^* = x_{i^*}$  is called a size-biased pick from  $(x_1, x_2, \dots)$ . Given  $i^*$ , we remove the  $i^*$ th entry from  $(x_1, x_2, \dots)$  and repeat the process. This yields a random re-ordering  $(x_1^*, x_2^*, \dots)$  of  $(x_1, x_2, \dots)$  called the size-biased order (if ever no positive terms remain, the rest of the size-biased elements are 0). Similarly for a random sequence  $(X_1, X_2, \dots)$  we define the size-biased ordering by first conditioning on the value of the sequence. For any finite measure  $\mu$  on  $\mathcal{S}^\downarrow$ , define the size-biased distribution  $\mu^*$  of  $\mu$  by

$$\mu^*(f) = \int_{\mathcal{S}^\downarrow} \mu(d\mathbf{s}) \mathbb{E}[f(\mathbf{s}^*)],$$

where  $\mathbf{s}^*$  is the size-biased reordering of  $\mathbf{s}$ .

Define the measure  $\nu_2$  on  $\mathcal{S}^\downarrow$  by

$$\int_{\mathcal{S}^\downarrow} \nu_2(d\mathbf{s}) f(\mathbf{s}) = \int_{1/2}^1 \sqrt{\frac{2}{\pi s_1^3 (1 - s_1)^3}} ds_1 f(s_1, 1 - s_1, 0, 0, \dots).$$

**Theorem 8.** *With the notation above,*

$$\lim_{n \rightarrow \infty} n^{1/2} (1 - s_1) \bar{q}_n(d\mathbf{s}) = \frac{\sigma_1 \sqrt{\xi(A)}}{2} (1 - s_1) \nu_2(d\mathbf{s}),$$

where the limit is taken in the sense of weak convergence of finite measures.



*Proof.* We follow the reductions in Section 5.1 of [7]. By Lemma 16 in [7] (which is a easy variation of Proposition 2.3 in [3]) it is sufficient to show that

$$\lim_{n \rightarrow \infty} n^{1/2}((1 - s_1)\bar{q}_n(d\mathbf{s}))^* = \frac{\sigma_1 \sqrt{\xi(A)}}{2}((1 - s_1)\nu_2(d\mathbf{s}))^*.$$

Note that for any finite non-negative measure  $\mu$  on  $\mathcal{S}^\downarrow$  and non-negative continuous function  $f : \mathcal{S}_1 \rightarrow \mathbb{R}$  we have

$$((1 - s_1)\mu(d\mathbf{s}))^*(f) = \int_{\mathcal{S}_1} \mu^*(d\mathbf{x})(1 - \max \mathbf{x})f(\mathbf{x}).$$

Consequently the theorem follows from the following Proposition.  $\square$

**Proposition 5.** *Let  $f : \mathcal{S}_1 \rightarrow \mathbb{R}$  be continuous and let  $g(\mathbf{x}) = (1 - \max \mathbf{x})f(\mathbf{x})$ . Then*

$$\sqrt{n}\bar{q}_n^*(g) \rightarrow \frac{\sigma_1 \sqrt{\xi(A)}}{\sqrt{2\pi}} \int_0^1 \frac{dx}{x^{1/2}(1-x)^{3/2}} g(x, 1-x, 0, \dots).$$

First note that, by linearity, we may assume that  $f \geq 0$  and  $\|f\|_\infty \leq 1$ . We begin the proof of this Proposition with several Lemmas regarding the concentration of mass of  $\bar{q}_n^*$ . We also note that for the remainder of this section we are following Section 5.1 in [7] very closely with minor differences to account for our more general setting and we invoke Corollary 2 rather than the Otter-Dwass formula – and we get different intermediate constants than they get, but the end results are the same. Nonetheless, the full computation is worth including because it makes clear why the factor of  $\sqrt{\xi(A)}$  appears in the scaling limit.

**Lemma 7.** *For every  $\epsilon > 0$ ,  $\sqrt{n}q_n(p(\lambda) > \epsilon\sqrt{n}) \rightarrow 0$  as  $n \rightarrow \infty$ .*

*Proof.* Observe that we have the local limit theorem (see e.g. Theorem 3.5.2 in [4])

$$(3.5) \quad \lim_{n \rightarrow \infty} \sup_{p \in \mathbb{Z}} \left| \sqrt{n}\mathbb{P}(S_n = -p) - \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{p^2}{2n\sigma^2}\right) \right| = 0.$$

Using this and Equation (3.4) we have  $q_n(p(\lambda) = p) \leq C\hat{\xi}(p)$  for some  $C$  independent of  $n$  and  $p$ . Since  $\xi$  has finite variance,  $\hat{\xi}$  has finite mean so  $\hat{\xi}((k, \infty)) = o(k^{-1})$ . The result follows.  $\square$

Note that an immediate consequence of the local limit theorem is that for any fixed  $k > 0$  we have

$$\lim_{n \rightarrow \infty} \sup_{1 \leq p \leq kn^{1/2}} \left| \sigma\sqrt{2\pi n} \exp\left(\frac{p^2}{2n\sigma^2}\right) \mathbb{P}(S_n = -p) - 1 \right| = 0,$$

and this is often the result we are really using when we cite the local limit theorem.

**Lemma 8.** *For  $g$  as in Proposition 5 we have*

$$\lim_{\eta \downarrow 0} \limsup_{n \rightarrow \infty} \sqrt{n}\bar{q}_n^*(|g|\mathbf{1}_{\{x_1 > 1-\eta\}}) = 0 \quad \text{and} \quad \lim_{n \rightarrow \infty} \sqrt{n}\bar{q}_n^*(\mathbf{1}_{\{x_1 < n^{-7/8}\}}) = 0.$$

*Proof.* Fix  $\eta > 0$ . Since we are assuming  $\|f\|_\infty \leq 1$ , we know that  $|g(\mathbf{x})| \leq (1 - x_1)$ . Observing that

$$\mathbb{P}(X_1^* = m | X_1 + \dots + X_p = n) = \frac{pm}{n} \frac{\mathbb{P}(X_1 = m) \mathbb{P}(X_2 + \dots + X_p = n - m)}{\mathbb{P}(X_1 + \dots + X_p = n)},$$

and using (ii) in Lemma 6 we see that  $\sqrt{n} \bar{q}_n^*(|g| \mathbf{1}_{\{x_1 > 1-\eta\}})$  is bounded above by  $o(1)$  plus

$$n^{1/2} \sum_{1 \leq p \leq n^{1/2}} q_n(p(\lambda) = p) \sum_{(1-\eta)n \leq m_1} \left(1 - \frac{m_1}{n - \mathbf{1}_p}\right) \frac{pm_1}{n - \mathbf{1}_p} \frac{\mathbb{P}(X_1 = m_1) \mathbb{P}(\tau_{p-1} = n - m_1 - \mathbf{1}_p)}{\mathbb{P}(\tau_p = n - \mathbf{1}_p)},$$

where the  $o(1)$  term is justified by Lemma 7 and our restriction to  $1 \leq p \leq n^{1/2}$ . Observe that Equation (3.5) implies that

$$\mathbb{P}(\tau_1 = n) = \frac{1}{n} \mathbb{P}(S_n = -1) \sim \frac{1}{\sigma \sqrt{2\pi}} n^{-3/2}.$$

Using Corollary 2, and again that  $q_n(p(\lambda) = p) \leq C \hat{\xi}(p)$  for some  $C$  independent of  $n$  and  $p$ , we find that for large  $n$ ,  $\sqrt{n} \bar{q}_n^*(|g| \mathbf{1}_{\{x_1 > 1-\eta\}})$  is bounded above by  $o(1)$  plus

$$\begin{aligned} & C n^{1/2} \sum_{1 \leq p \leq n^{1/2}} p^2 \xi(p) \\ & \times \sum_{(1-\eta)n \leq m_1 < n - \mathbf{1}_p} \left(1 - \frac{m_1}{n - \mathbf{1}_p}\right) \frac{m_1}{n - \mathbf{1}_p} m_1^{-3/2} \frac{n - \mathbf{1}_p}{n - m_1 - \mathbf{1}_p} \frac{\mathbb{P}(S_{n-m_1-\mathbf{1}_p} = -p + 1)}{\mathbb{P}(S_{n-\mathbf{1}_p} = -p)}. \end{aligned}$$

Simplifying we get

$$\sqrt{n} \bar{q}_n^*(|g| \mathbf{1}_{\{x_1 > 1-\eta\}}) \leq o(1) + C \sum_{1 \leq p \leq n^{1/2}} p^2 \xi(p) \frac{1}{n} \sum_{(1-\eta)n \leq m_1 \leq n - \mathbf{1}_p} \sqrt{\frac{n}{m_1}} \frac{\mathbb{P}(S_{n-m_1-\mathbf{1}_p} = -p + 1)}{\mathbb{P}(S_{n-\mathbf{1}_p} = -p)}.$$

Equation (3.5) implies that, for  $1 \leq p \leq n^{1/2}$ ,  $\sqrt{n} \mathbb{P}(S_{n-\mathbf{1}_p} = -p)$  and  $\sqrt{n - m_1} \mathbb{P}(S_{n-\mathbf{1}_p} = -p + 1)$  are bounded below and above respectively for some constants independent of  $n$  and  $p$ . Hence we have

$$\begin{aligned} \sqrt{n} \bar{q}_n^*(|g| \mathbf{1}_{\{x_1 > 1-\eta\}}) & \leq o(1) + C \sum_{1 \leq p \leq n^{1/2}} p^2 \xi(p) \frac{1}{n} \sum_{(1-\eta)n \leq m_1 < n - \mathbf{1}_p} \frac{1}{\sqrt{\frac{m_1}{n} \left(1 - \frac{m_1}{n}\right)}} \\ & \leq o(1) + C \sum_{p=1}^{\infty} p^2 \xi(p) \frac{1}{n} \sum_{(1-\eta)n \leq m_1 < n - \mathbf{1}_p} \frac{1}{\sqrt{\frac{m_1}{n} \left(1 - \frac{m_1}{n}\right)}}. \end{aligned}$$

Note that the  $m_1 = n$  term has been absorbed into the  $o(1)$  term. The upper bound converges to  $C \int_{1-\eta}^1 (x(1-x))^{-1/2} dx$ , which goes to 0 and  $\eta \rightarrow 0$ . A little bit of care must be taken here since the integral is improper as a Riemann integral, however this is fine since the sums actually under approximate the integral in this case.

The second limit can be proved in a similar fashion. Note that  $\sqrt{n}\bar{q}_n^*(\mathbf{1}_{\{x_1 < n^{-7/8}\}})$  is bounded above by

$$\begin{aligned}
& n^{1/2} \sum_{1 \leq p \leq n^{1/2}} q_n(p(\lambda) = p) \sum_{m_1 \leq n^{1/8}} \frac{pm_1}{n - \mathbf{1}_p} \frac{\mathbb{P}(X_1 = m_1)\mathbb{P}(\tau_{p-1} = n - m_1 - \mathbf{1}_p)}{\mathbb{P}(\tau_p = n - \mathbf{1}_p)} + o(1) \\
& \leq Cn^{-3/8} \sum_{1 \leq p \leq n^{1/2}} p^2 \xi(p) \sum_{1 \leq m_1 \leq n^{1/8}} \frac{\mathbb{P}(\tau_{p-1} = n - m_1 - \mathbf{1}_p)}{\mathbb{P}(\tau_p = n - \mathbf{1}_p)} + o(1) \\
& \leq Cn^{-3/8} \sum_{1 \leq p \leq n^{1/2}} p^2 \xi(p) \sum_{1 \leq m_1 \leq n^{1/8}} \frac{\mathbb{P}(S_{n-m_1-\mathbf{1}_p} = -p+1)}{\mathbb{P}(S_{n-\mathbf{1}_p} = -p)} + o(1) \\
& \leq Cn^{-1/4} \sum_{1 \leq p \leq n^{1/2}} p^2 \xi(p) + o(1),
\end{aligned}$$

where the last step is justified by the local limit theorem.  $\square$

**Lemma 9.** *For every  $\eta > 0$  we have*

$$\lim_{n \rightarrow \infty} \sqrt{n}\bar{q}_n^*(\mathbf{1}_{\{x_1+x_2 < 1-\eta\}}) = 0.$$

*Proof.* Fix  $0 < \epsilon < 1$ . Up to addition by an  $o(1)$  term depending on  $\epsilon$  we have that  $\sqrt{n}\bar{q}_n^*(\mathbf{1}_{\{x_1+x_2 < 1-\eta\}})$  is bounded above by

$$\begin{aligned}
& C\sqrt{n} \sum_{1 \leq p \leq \epsilon n^{1/2}} p \xi(p) \\
& \times \sum_{m_1+m_2 \leq (1-\eta)n} \frac{pm_1}{n - \mathbf{1}_p} \frac{(p-1)m_2}{n - m_1 - \mathbf{1}_p} \frac{\mathbb{P}(X_1 = m_1)\mathbb{P}(X_2 = m_2)\mathbb{P}(\tau_{p-2} = n - m_1 - m_2 - \mathbf{1}_p)}{\mathbb{P}(\tau_p = n - \mathbf{1}_p)},
\end{aligned}$$

where  $m_1, m_2 \geq 1$ . If  $m_1 + m_2 \leq (1-\eta)n$  then  $n - m_1 - m_2 \geq \eta n$  and, in particular, the quantity on the left goes to  $\infty$  as  $n$  does. Consequently Corollary 2 and Equation (3.5) imply that

$$\frac{\mathbb{P}(\tau_{p-2} = n - m_1 - m_2 - \mathbf{1}_p)}{\mathbb{P}(\tau_p = n - \mathbf{1}_p)} \leq \frac{C}{\eta^{3/2}},$$

for  $C$  independent of  $1 \leq p \leq \epsilon n^{1/2}$ . Our assumption that  $\epsilon < 1$  implies that  $C$  is independent of  $\epsilon$  as well. Again using that  $n^{3/2}\mathbb{P}(X_1 = n)$  is bounded we have

$$\begin{aligned}
\sqrt{n}\bar{q}_n^*(\mathbf{1}_{\{x_1+x_2 < 1-\eta\}}) & \leq o(1) + \frac{C}{\eta^{5/2}\sqrt{n}} \sum_{1 \leq p \leq \epsilon n^{1/2}} p^3 \xi(p) \frac{1}{n^2} \sum_{m_1+m_2 \leq (1-\eta)n} \sqrt{\frac{n}{m_1}} \sqrt{\frac{n}{m_2}} \\
& \leq o(1) + \frac{C\epsilon}{\eta^{5/2}} \sum_{p=1}^{\infty} p^2 \xi(p) \int_0^1 \int_0^1 \frac{dx dy}{\sqrt{xy}}.
\end{aligned}$$

Taking the lim sup as  $n \rightarrow \infty$  and then letting  $\epsilon \rightarrow 0$  yields the result.  $\square$

**Lemma 10.** *There exists a function  $\beta_\eta = o(\eta)$  as  $\eta \downarrow 0$  such that*

$$\begin{aligned} \lim_{\eta \downarrow 0} \liminf_{n \rightarrow \infty} \sqrt{n} \bar{q}_n^*(g \mathbf{1}_{\{x_1 < 1-\eta, x_1+x_2 > 1-\beta_\eta\}}) &= \lim_{\eta \downarrow 0} \limsup_{n \rightarrow \infty} \sqrt{n} \bar{q}_n^*(g \mathbf{1}_{\{x_1 < 1-\eta, x_1+x_2 > 1-\beta_\eta\}}) \\ &= \frac{\sigma_1 \sqrt{\xi(A)}}{\sqrt{2\pi}} \int_0^1 \frac{g((x, 1-x, 0, \dots))}{x^{1/2}(1-x)^{3/2}} dx. \end{aligned}$$

*Proof.* Fix  $\eta > 0$  and suppose that  $\eta' \in (0, \eta)$ . Using Lemmas 7 and 8 we decompose according to the events  $\{p(\lambda) > \epsilon \sqrt{n}\}$  and  $\{\mathbf{x} : x_1 \leq n^{-7/8}\}$  to get

$$\begin{aligned} (3.6) \quad \sqrt{n} \bar{q}_n^*(g \mathbf{1}_{\{x_1 < 1-\eta, x_1+x_2 > 1-\eta'\}}) &= o(1) + \sqrt{n} \sum_{1 \leq p \leq \epsilon n^{1/2}} q_n(p(\lambda) = p) \\ &\times \sum_{\substack{n^{1/8} \leq m_1 \leq (1-\eta)(n-\mathbf{1}_p) \\ (1-\eta')(n-\mathbf{1}_p) \leq m_1+m_2 \leq n-\mathbf{1}_p}} \mathbb{E}[g((m_1, m_2, X_3^*, \dots, X_p^*, 0, \dots)/(n-\mathbf{1}_p)) | \tau_p = n-\mathbf{1}_p, X_1^* = m_1, X_2^* = m_2] \\ &\times \frac{pm_1}{n-\mathbf{1}_p} \frac{\frac{(p-1)m_2}{n-\mathbf{1}_p}}{1 - \frac{m_1}{n-\mathbf{1}_p}} \mathbb{P}(X_1 = m_1) \mathbb{P}(X_2 = m_2) \frac{\mathbb{P}(\tau_{p-2} = n - m_1 - m_2 - \mathbf{1}_p)}{\mathbb{P}(\tau_p = n - \mathbf{1}_p)}. \end{aligned}$$

Observe that, if  $1 \geq x_1 + x_2 \geq 1 - \eta'$  and  $x_1 \leq 1 - \eta$ , then  $x_2/(1-x_1) \geq 1 - \eta'/\eta$  and  $(1-x_1)/x_2 \geq 1$ .

Using the local limit theorem and we observe that

$$\sup_{1 \leq p \leq \epsilon n^{1/2}} \left| \sigma \sqrt{2\pi n} \exp\left(\frac{p^2}{2n\sigma^2}\right) \mathbb{P}(S_n = -p) - \sigma \sqrt{2\pi n} \exp\left(\frac{1}{2n\sigma^2}\right) \mathbb{P}(S_n = -1) \right| \rightarrow 0$$

Consequently,

$$\sup_{1 \leq p \leq \epsilon n^{1/2}} \left| \frac{\frac{q_n(p(\lambda)=p)}{\hat{\xi}(p)}}{\exp\left(\frac{1-p^2}{2n\sigma^2}\right)} - 1 \right| = \sup_{1 \leq p \leq n^{1/2}} \left| \frac{\sqrt{2\pi\sigma^2 n} \exp(p^2/2n\sigma)(n-\mathbf{1}_p) \mathbb{P}(S_n = -p)}{\sqrt{2\pi\sigma^2 n} \exp(1/2n\sigma) n \mathbb{P}(S_n = -1)} - 1 \right| \rightarrow 0.$$

Thus, for sufficiently large  $n$  and small  $\epsilon$ , we have

$$1 - \eta \leq \frac{q_n(p(\lambda) = p)}{\hat{\xi}(p)} \leq 1 + \eta,$$

for all  $1 \leq p \leq \epsilon n^{1/2}$ . Using the local limit theorem and Corollary 2 we have

$$\sup_{1 \leq p \leq \epsilon n^{1/2}} \left| p^{-1} \sigma \sqrt{2\pi} n^{3/2} \exp\left(\frac{p^2}{2n\sigma^2}\right) \mathbb{P}(\tau_p = n) - 1 \right| \rightarrow 0.$$

Thus, for sufficiently large  $n$  and small  $\epsilon$ , we have

$$\frac{1-\eta}{\sigma \sqrt{2\pi}} \leq p^{-1} n^{3/2} \mathbb{P}(\tau_p = n) \leq \frac{1+\eta}{\sigma \sqrt{2\pi}},$$

for all  $1 \leq p \leq \epsilon n^{1/2}$ . We note in particular that  $\tau_1 = X_1 =_d X_2$ . Furthermore, for  $n^{1/8} \leq m_1 \leq (1-\eta)n$  and  $m_1 + m_2 \geq (1-\eta')n$  we have  $m_2 \geq (\eta - \eta')n$  so that  $m_1$  and  $m_2$

go to infinity as  $n$  does. Thus, for large  $n$  (how large now depends on  $\eta'$ ) we have

$$\frac{(1-\eta)^2}{2\pi\sigma^2} \leq (m_1 m_2)^{3/2} \mathbb{P}(X_1 = m_1) \mathbb{P}(X_2 = m_2) \leq \frac{(1+\eta)^2}{2\pi\sigma^2}.$$

Now, recall that  $f$  is uniformly continuous on  $\mathcal{S}_1$ . Furthermore, on the set  $\{\mathbf{x} \in \mathcal{S}_1 : x_1 + x_2 > 3/4\}$  we have  $\max \mathbf{x} = x_1 \vee x_2$  and  $\mathbf{x} \mapsto \max \mathbf{x}$  is thus uniformly continuous on this set. Therefore for  $\eta' < (1/4) \wedge \eta^2$  sufficiently small we have

$$|g((m_1, m_2, m_3, \dots)/n) - g((m_1, n - m_1, 0, \dots)/n)| \leq \eta,$$

for every  $(m_1, m_2, \dots)$  with sum  $n$  sufficiently large such that  $m_1 + m_2 \geq (1 - \eta')n$ . Take  $\beta_\eta := \eta'$ .

Given the symmetry of the bounds we have just established it is easy to see that the proofs for the lim sup and lim inf will be nearly identical, one using the upper bounds and the other the lower. We will only write down the proof for the lim inf. For sufficiently large  $n$  we have that, up to addition of an  $o(1)$  term,  $\sqrt{n} \bar{q}_n^*(g \mathbf{1}_{\{x_1 < 1-\eta, x_1+x_2 > 1-\eta'\}})$  is bounded below by

$$\begin{aligned} & \frac{(1-\eta)^3(1-\eta'/\eta)}{(1+\eta)} \sum_{1 \leq p \leq \epsilon n^{1/2}} (p-1) \hat{\xi}(p) \frac{1}{n - \mathbf{1}_p} \\ & \times \sum_{n^{1/8} \leq m_1 \leq (1-\eta)(n - \mathbf{1}_p)} (g((m_1, n - m_1 - \mathbf{1}_p, 0, \dots)/(n - \mathbf{1}_p)) - \eta) \\ & \times \frac{1}{(m_1/(n - \mathbf{1}_p))^{1/2}} \frac{1}{(1 - m_1/(n - \mathbf{1}_p))^{3/2}} \frac{1}{\sigma \sqrt{2\pi}} \\ & \times \sum_{(1-\eta')(n - \mathbf{1}_p) - m_1 \leq m_2 \leq n - m_1 - \mathbf{1}_p} \mathbb{P}(\tau_{p-2} = n - m_1 - m_2 - \mathbf{1}_p). \end{aligned}$$

Observe that this last sum is equal to  $\sum_{m=0}^{\eta'(n - \mathbf{1}_p)} \mathbb{P}(\tau_{p-2} = m)$ . By the local limit theorem, this can be made arbitrarily close to 1 independent of  $1 \leq p \leq n^{1/2}$ . Using the convergence of Riemann sums (again care must be taken since the integral we get is improper), we have

$$\begin{aligned} & \liminf_{n \rightarrow \infty} \sqrt{n} \bar{q}_n^*(g \mathbf{1}_{\{x_1 < 1-\eta, x_1+x_2 > 1-\eta'\}}) \\ & \geq \frac{(1-\eta)^3(1-\eta'/\eta)}{1+\eta} \sum_{p=1}^{\infty} (p-1) \hat{\xi}(p) \int_0^{1-\eta} \frac{dx}{\sigma \sqrt{2\pi} x^{1/2} (1-x)^{3/2}} (g(x, 1-x, 0, \dots) - \eta) \end{aligned}$$

Letting  $\eta \downarrow 0$  coupled with observing that  $\sum_{p=1}^{\infty} (p-1) \hat{\xi}(p) = \sigma_1^2$  and recalling that  $\sigma^2 = \sigma_1^2/\xi(A)$  completes the proof.  $\square$

*Proof of Proposition 5.* Observe that

$$|\bar{q}_n^*(g) - \bar{q}_n^*(g \mathbf{1}_{\{x_1 < 1-\eta, x_1+x_2 > 1-\eta'\}})| \leq \bar{q}_n^*(|g| \mathbf{1}_{\{x_1 \geq 1-\eta\}}) + \bar{q}_n^*(|g| \mathbf{1}_{\{x_1+x_2 \leq 1-\eta'\}}).$$

Fix  $\epsilon > 0$  and apply Lemmas 8 and 10 to find  $\eta, \eta'$  such that

$$\sqrt{n} \bar{q}_n^*(|g| \mathbf{1}_{\{x_1 \geq 1-\eta\}}) < \frac{\epsilon}{2}$$

and

$$\left| \sqrt{n} \bar{q}_n^*(g \mathbf{1}_{\{x_1 < 1-\eta, x_1+x_2 > 1-\beta_\eta\}}) - \frac{\sigma_1 \sqrt{\xi(A)}}{\sqrt{2\pi}} \int_0^1 \frac{g(x, 1-x, 0, 0, \dots)}{x^{1/2}(1-x)^{3/2}} dx \right| \leq \frac{\epsilon}{2},$$

for large enough  $n$ . For this choice of  $\eta, \eta'$  and large  $n$  we have

$$\left| \bar{q}_n^*(g) - \frac{\sigma_1 \sqrt{\xi(A)}}{\sqrt{2\pi}} \int_0^1 \frac{g(x, 1-x, 0, 0, \dots)}{x^{1/2}(1-x)^{3/2}} dx \right| \leq \epsilon + \bar{q}_n^*(|g| \mathbf{1}_{\{x_1+x_2 \leq 1-\eta'\}}).$$

By Lemma 9 the upper bound goes to  $\epsilon$  as  $n \rightarrow \infty$ , and the result follows.  $\square$

As an immediate corollary of these results, we also identify the unnormalized limit of  $\bar{q}_n$ .

**Corollary 3.**  $\bar{q}_n \xrightarrow{d} \delta_{(1,0,0,\dots)}$ .

*Proof.* Taking  $f \equiv 1$  in Proposition 5 gives  $\bar{q}_n(1-s_1) \rightarrow 0$ . Thus  $\bar{q}_n(s_1) \rightarrow 1$ . Suppose that there exists  $0 < \eta < 1$  such that  $\bar{q}_n(s_1 < \eta) \not\rightarrow 0$ . Let  $\epsilon = \limsup_{n \rightarrow \infty} \bar{q}_n(\mathbf{1}_{\{s_1 < \eta\}}) > 0$ . Observing that

$$\begin{aligned} \bar{q}_n(s_1) &= \bar{q}_n(s_1 \mathbf{1}_{\{s_1 < \eta\}}) + \bar{q}_n(s_1 \mathbf{1}_{\{s_1 \geq \eta\}}) \\ &\leq 1 - (1-\eta) \bar{q}_n(s_1 < \eta), \end{aligned}$$

we see that

$$\liminf_{n \rightarrow \infty} \bar{q}_n(s_1) \leq 1 - (1-\eta)\epsilon < 1,$$

a contradiction. Thus  $\bar{q}_n(\mathbf{1}_{\{s_1 < \eta\}}) \rightarrow 0$  for all  $0 < \eta < 1$  and, consequently,  $\bar{q}_n(s_1 \geq \eta) \rightarrow 1$ .  $\square$

Note that, as a consequence of Equation (3.4), we have  $q_n(p(\lambda) = p) \rightarrow \hat{\xi}(p)$ . Thus, while the degree of the root vertex may be large, only one of the trees attached to the root will have noticeable size.

**3.3. Convergence of Galton-Watson trees.** We are now prepared to prove Theorem 1, which, after all of our work above, is a rather straight forward.

*Proof.* Lemma 6 shows that  $T_n^A$  has law  $\mathbf{P}_n^q$  for a particular choice of  $(q_n)_{n \geq 1}$ . Theorem 8 then shows that the hypotheses of Theorem 5 are satisfied.  $\square$

## REFERENCES

- [1] David Aldous. The continuum random tree. III. *Ann. Probab.*, 21(1):248–289, 1993.
- [2] Jean Bertoin. The asymptotic behavior of fragmentation processes. *J. Eur. Math. Soc. (JEMS)*, 5(4):395–416, 2003.
- [3] Jean Bertoin. *Random fragmentation and coagulation processes*, volume 102 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2006.
- [4] Rick Durrett. *Probability: theory and examples*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, fourth edition, 2010.
- [5] Bénédicte Haas. Loss of mass in deterministic and random fragmentations. *Stochastic Process. Appl.*, 106(2):245–277, 2003.
- [6] Bénédicte Haas and Grégory Miermont. The genealogy of self-similar fragmentations with negative index as a continuum random tree. *Electron. J. Probab.*, 9:no. 4, 57–97 (electronic), 2004.

- [7] Bénédicte Haas and Grégory Miermont. Scaling limits of Markov branching trees, with applications to Galton-Watson and random unordered trees. *arXiv:1003.3632v2 [math.PR]*, 2010.
- [8] Bénédicte Haas, Grégory Miermont, Jim Pitman, and Matthias Winkel. Continuum tree asymptotics of discrete fragmentations and applications to phylogenetic models. *Ann. Probab.*, 36(5):1790–1837, 2008.
- [9] Jean-François Le Gall. Random real trees. *Ann. Fac. Sci. Toulouse Math. (6)*, 15(1):35–62, 2006.
- [10] Jean-François Le Gall. Itô’s excursion theory and random trees. *Stochastic Process. Appl.*, 120(5):721–749, 2010.
- [11] Nariyuki Minami. On the number of vertices with a given degree in a Galton-Watson tree. *Adv. in Appl. Probab.*, 37(1):229–264, 2005.
- [12] J. Pitman. *Combinatorial stochastic processes*, volume 1875 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 2006. Lectures from the 32nd Summer School on Probability Theory held in Saint-Flour, July 7–24, 2002.

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF CALIFORNIA AT BERKELEY, BERKELEY, CA 94720  
E-mail address: drizzolo@math.berkeley.edu